

DOCUMENT RESUME

ED 470 589

TM 034 565

AUTHOR Mertler, Craig A.
TITLE Using Standardized Test Data To Guide Instruction and Intervention. ERIC Digest.
INSTITUTION ERIC Clearinghouse on Assessment and Evaluation, Washington, DC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO EDO-TM-02-07
PUB DATE 2002-10-00
NOTE 4p.
CONTRACT ED-99-CO-0032
AVAILABLE FROM ERIC Clearinghouse on Assessment and Evaluation, 1129 Shriver Laboratory, University of Maryland, College Park, MD 20742. Tel: 800-464-3742 (Toll Free).
PUB TYPE ERIC Publications (071) -- ERIC Digests in Full Text (073)
EDRS PRICE EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Intervention; Research Reports; *Standardized Tests; *Teacher Role; Test Interpretation; Test Results; *Test Use
IDENTIFIERS ERIC Digests

ABSTRACT

This Digest addresses two ways that classroom teachers can use the results of standardized tests: (1) to revise instruction for entire classes or courses; and (2) to develop specific intervention strategies for individual students. Test publishing companies typically provide classroom-level reports to enable teachers to see how a group of students performs. Teachers should examine such reports to see areas in which high percentages of children are performing below average. They then must consider whether the curriculum and the test are aligned and decide what they can do to remedy the situation. An example report from Riverside, the company that publishes the Iowa Tests of Basic Skills, illustrates the kind of data the test published may provide and how to read and use it. Standardized test data can also be used to guide the development of individualized intervention strategies. Nearly all publishers of standardized achievement tests provide both criterion- and norm-referenced results on individual student reports. Bearing in mind that a given subtest may consist of only a few items, the teacher can identify areas in which the student performed below the average, and can develop appropriate interventions. Teachers may wish to begin their inquiry into test results by focusing on such scores as national percentile ranks and their associated confidence bands. Limiting the data to be interpreted and understanding what scores really mean make the process more efficient. (SLD)

Using Standardized Test Data to Guide Instruction and Intervention

Craig A. Mertler

Bowling Green (OH) State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

ERIC Digest EDO-TM-02-07

October 2002

BEST COPY AVAILABLE

Using Standardized Test Data To Guide Instruction and Intervention

Craig A. Mertler
Bowling Green (OH) State University

When teachers review test score reports, they may find the sheer volume of information presented overwhelming, and they may also be unsure how to interpret and use results in the classroom. While the idea of data-driven decision-making is not new, it does require a special skill to focus on a few key pieces of information from a test and use them to make instructional changes. This Digest addresses two ways that classroom teachers can use the results of standardized tests: (1) to revise instruction for entire classes or courses and (2) to develop specific intervention strategies for individual students.

Using Test Scores To Revise Group Instruction

Test publishing companies typically provide classroom-level reports to enable teachers to see how a group of students performs across the curriculum. Even if a group of students has moved on by the time score reports are available, teachers should examine class-level results as a source of information for revising curriculum and instruction for the next class. Content areas or subtests in which high percentages of children are performing below average indicate areas of deficiency.

Once teachers have noted and prioritized deficiencies, they may consider one or more of the following questions:

- Where is this content addressed in our district's curriculum?
- At what point in the school are these concepts/skills taught?
- How are the students taught these concepts/skills?
- How are students required to demonstrate that they have mastered the concepts/skills? In other words, how are they assessed in the classroom?

Answers to these questions should point the way to new methods of instruction, reinforcement, or assessment (Mertler, 2001, 2003). They may also introduce evidence that the curriculum and the tests are not in alignment.

Revising Group Instruction: An Example

While reports from state tests and tests from commercial publishers vary in format, most feature certain common elements. Riverside, which publishes the Iowa Tests of Basic Skills, provides an illustrative sample class performance profile at the following Web address: http://www.riverpub.com/products/group/itbs_a/scoring.html#grpperm

This report, like many others, offers both *norm-referenced test results*, which allow performance comparisons with other groups of students taking the test, and *criterion-referenced information*, which provides data such as how many questions students attempted

and how many correct answers they gave for each category of question. Language skills might, for example, involve subtests of spelling, capitalization, punctuation, and usage; mathematics might break down to concepts, problem solving, data interpretation, and computation. Sometimes you'll also be able to see the number of questions devoted to each area within a subtest (e.g., in the area of mathematics concepts, how many questions deal with number properties, with algebra, with geometry, with measurement, and with estimation).

Typical scores reported might also include the following:

Standard (or "Scale") score (SS): A score that has been transformed mathematically and put on a scale to allow comparisons with different forms and levels of a test.

Grade equivalent (GE) Score: A norm-referenced score that indicates the grade and month of the school year for which a score is average. The average score for a fifth grader being tested in the seventh month of the school year would be 5.7. If a child has a GE score well above his or her grade in school—a fifth grader with a GE of 9.1 on a reading subtest, for example—it doesn't mean that the child can do ninth-grade work, but rather, that he or she scored the same as an average entering ninth grader would if the ninth grader took the fifth-grade test.

National percentile rank (NPR): the percentage of students in the norm group that performed at or below a particular performance level. It's important to note the group to which students are being compared. Some test publishers provide separate norms for, say, large urban school districts across the country, or Catholic schools, while also providing norms based on a representative sample of test-takers across the country, and/or other groups taking the test in the state.

Normal curve equivalent (NCE): A normalized standardized score with a mean of 50 and a standard deviation of 21.06 resulting in a near equal interval scale from 0 to 99. The NCE was developed by RMC Research Corporation in 1976 to measure the effectiveness of the Title I Program across the United States and is often used to measure gains over time.

National stanine (NS): Stanine scores range from 1 to 9, with a score of 5 representing an average range. The percentage of scores at each stanine level in a normalized standard score scale is 4, 7, 12, 17, 20, 17, 12, 7, and 4, respectively. Percentile rank scores provide similar, though more precise, information. For example, a percentile rank near the middle of the distribution (e.g., 45 to 55) will be roughly equivalent to a stanine score of 5.

A class test report might also present a graphic illustrating *confidence bands*, which represent the margins of error for individual subtests. Studying them will permit a teacher to get a quick overview of a class' performance because non-overlapping bands indicate that scores are truly or significantly different from each other. For example, the students in a class might perform significantly lower on "Vocabulary" than on "Reading Comprehension."



It is helpful to identify the subtest(s) upon which a particular class achieves at a national percentile rank of below 50 in order to make these content areas targets for possible instructional change. Alternatively, teachers might look at the skill areas in which high percentages of students scored in the bottom 25 percent or low percentages of students scored in the top 25 percent. Again, teacher would want to rank order, or otherwise prioritize, these areas for possible revision of instruction.

Using Test Scores To Design Individualized Interventions

Standardized test data may also be used very effectively in order to guide the development of individualized intervention strategies. First, however, it is important to remember that general achievement tests are intended to survey basic skills across a broad domain of content (Chase, 1999). On almost any standardized achievement test, a given subtest may consist of as few as five or six items. The fewer the number of items on a subtest, the less reliable the scores will be (Airasian, 2000, 2001). Careless errors or lucky guesses by students may substantially alter the score on that subtest, especially if scores are reported as percentages of items answered correctly or as percentile ranks. Therefore, it is important not only to examine the raw scores and percentile ranks, but also the total number of items possible on a given test prior to making any intervention decisions (Mertler, 2003).

Nearly all publishers of standardized achievement tests provide both criterion- and norm-referenced results on individual student reports. Many results are reported in terms of *average* performance (i.e., below average, average, above average). It is again important to remember that "average" simply means that half of the norm group scored above and half scored below that particular score (Gallagher, 1998). Teachers should take great care to avoid the *overinterpretation* of test scores (Airasian, 2000, 2001).

The process for examining test results in order to help guide the development of intervention strategies for individual students is essentially the same as for the whole class. First, the teacher identifies any content areas or subtests in which the student performed below average. Second, the teacher establishes priorities among these areas, selecting a workable number of content areas, perhaps one or two, to serve as the focus of an intervention. Third, the teacher identifies new or different resource materials, methods of instruction, reinforcement, and/or assessment in order to meet the needs of the individual student. The success of this intervention may be monitored both through classroom assessments and on future test scores. Given the length of time it takes for test scores to become available, it may be that a teacher at the next grade level will have to follow through on the intervention.

Designing Intervention Strategies: An Example

Individual score reports, like classroom performance reports, tend to contain both norm- and criterion-referenced information. The results may include scaled scores, grade-equivalent scores, national stanines, normal curve equivalent scores, national percentiles, and national percentile bands as defined above. The national percentile ranks and associated confidence bands allow the teacher to see how well a particular student performed in relation to the national norm group on the various subtests. An individual skill report might also include a breakdown of how many items the test included in each skill area, how many the student attempted, how many he or she got correct, and how this performance compared to the national norm

group.

By way of illustration, readers may wish to view a score report for a fictitious student, Mary Sanders, who took the *Iowa Tests of Basic Skills*. This score report (http://www.riverpub.com/products/group/itbs_a/scoring.html#indperm) shows percentile rank scores ranging from a low of 20 ("Capitalization") to a high of 71 ("Reading Comprehension" and "Listening"), with the student's performance substantially below the norm group in several areas, including "Capitalization," "Word Analysis," "Vocabulary," "Math Concepts and Estimation," and "Math Computation." These are potential areas the classroom teacher would want to target for possible interventions. Further examination of the "Math Concepts and Estimation" portion in the criterion-referenced section pinpoint Mary's difficulties to the items that dealt with "Algebra." Other areas in which Mary scored in the "Low" category include "Punctuation: Commas," "Math: Problem Solving: Approaches and Procedures," and "Social Studies: Economics." Again, this information would likely prove most essential in designing an intervention plan for Mary. Teachers can perform similar analyses on individual score reports from their own state tests or other commercial tests. The important point to remember is to use the data to identify specific areas of difficulty in order to plan a well-targeted intervention.

Conclusion

Teachers can learn to use empirical test data to assist in instructional decision-making for their classes or for individual students. To avoid being overwhelmed by data, especially since much of the information provided on test reports is analogous, teachers might wish to begin their inquiry by focusing on such scores as national percentile ranks and their associated confidence bands. Interpreting standardized test data for use in making instructional decisions *does* take some practice. Limiting the data to be interpreted—and understanding what those scores really mean—makes the process more efficient and allows teachers to make valuable use of their students' standardized test data to bring about increased achievement.

References

- Airasian, P. W. (2000). *Assessment in the classroom: A concise approach* (2nd ed.). Boston: McGraw-Hill.
- Airasian, P. W. (2001). *Classroom assessment: Concepts and applications* (4th ed.). Boston: McGraw-Hill.
- Chase, C. I. (1999). *Contemporary assessment for educators*. New York: Longman.
- Gallagher, J. D. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Merrill.
- Iowa Tests of Basic Skills—Individual Performance Profile*. (2002). Retrieved July 9, 2002, from http://www.riverpub.com/products/group/itbs_a/scoring.html#indperm
- Mertler, C. A. (2001). *Interpreting proficiency test data: Guiding instruction and intervention*. Unpublished manuscript (inservice training materials), Bowling Green State University.
- Mertler, C. A. (2003). *Classroom assessment: A practical guide for educators*. Los Angeles, CA: Pyrczak.



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)



NOTICE

Reproduction Basis

- ☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
- ☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").